Prosody Interaction with Speech Planning and Comprehension, a Survey of Interesting Effects in the Literature

David Porteš

Faculty of Informatics, Masaryk University,
Botanická 68a, 602 00 Brno, Czech Republic
xportes@fi.muni.cz

Abstract. Prosody is often treated as a peripheral, emotional embellishment of the linguistic signal, yet an increasing body of work reveals that it is involved in many aspects of language processing. This paper surveys multiple lines of evidence. First, research on speech planning shows that speakers do not merely assemble syntax and then append word-forms, but instead they construct a prosodic skeleton before lexical material is fully specified. Second, comprehension studies demonstrate that listeners exploit prosodic cues to resolve syntactic ambiguities well before the utterance is complete, and that even silent readers project a default prosodic contour onto text, influencing ambiguity resolution in the absence of any acoustic signal. Interestingly, the same sentence can elicit different prosodic contours in different languages, as evidenced by research on the Slavonic language of Croatian, among others. Together, these findings suggest that prosodic structure should be treated more as a fundamental, rather than complemental feature of language.

Keywords: Prosody; Speech Production; Speech Comprehension; Implicit Prosody; Prosodic Morphology

1 Introduction

In this paper, we describe various interesting findings in the literature that deal with prosodic effects and their role in speech planning and comprehension. While prosody is often relegated to mere complementary information source, which carries information on speaker's emotional state, these findings suggests that it in fact permeates all aspects of language in a very significant way. We draw on multiple branches of research, first being work on speech planning, which is concerned with finding the algorithm that turns an intention into a sentence, and then to concrete articulatory programs. Here we discuss how, while initial research started with a prosody last approach, important issues started to emerge that are making the field gravitate towards promoting prosody to a much bigger role. In the next section, we will show that not only in speech planning, but also in speech comprehension, there is evidence of prosody playing a much more important role than usually assumed. We will discuss the

A. Horák, P. Rychlý, A. Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2025, pp. 99–106, 2025.

results obtained by Beach[1], who has shown that prosodic cues help listeners disambiguate sentences even at a point where most of the sentence has not been uttered yet. Additionally, we will turn to a fascinating branch of research called the Implicit Prosody Hypothesis [3], which posits that, even in silent speech, where prosody is not present, readers are actually projecting prosody onto the text, as evidenced by prosodic effects influencing its interpretation.

2 Speech Planning

In this section, we follow a line of work on speech planning, which started by assuming that when speakers plan their sentences, they start by planning the surface syntax [7]. The word forms are then encoded in a left to right fashion, which makes it possible to start encoding word forms while only a part of the syntax is generated, and aligns with the notion that speakers often do not know how a sentence will end when they begin to utter it. However, from this initial view, a branch of research argues that planning has to be done in prosodic units, like the prosodic word or intonational phrase, rather than purely syntactical units, resulting in the Prosodic Hierarchy approach [5]. This approach argues that speakers must construct the whole prosodic structure of the sentence at the beginning, to explain certain effects that require look-ahead to future word forms. As an example, errors such as *chee kains* instead of *key chains* demonstrate

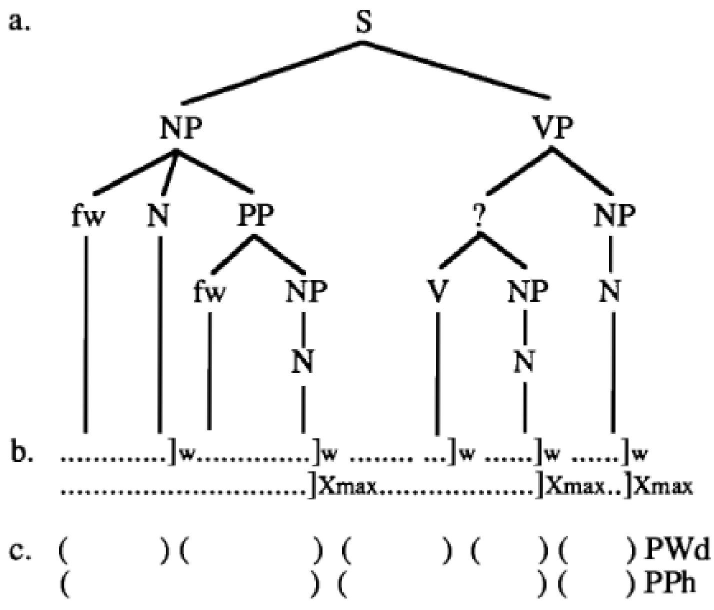

Fig. 1: A schema depicting how prosodic hierarchy is built from syntax. Adapted from [4]

that the speaker is aware of the following word [5]. While this example occurs in two adjacent words, there are other examples such as *Sicks and the Kneltics* for *Knicks and the Celtics*, or *you getter stop for bas* for *you better stop for gas*, where words further away are involved. These examples are lexical in nature, however, there are other look-ahead effects that operate on the stresses of upcoming words. The word *Japanese* has stress on the last syllable, while the word *institute* has stress on the initial one. When put together, as in **JapanESE INstitute**, the stress becomes **JAPanese INstitute**, to avoid a stress clash. While this could be explained by a limited look-ahead buffer, the fact that this effect can be iterated makes this explanation unlikely, such as in the following sentence:

(1) **MISSissippi LEGislators' ANtique NINEteen TWENty-seven MOtor-cars**.

Because of these look-ahead effects, the Prosodic Hierarchy proponents argue that speakers must construct a prosodic representation of the whole sentence first, based on syntax, so that they can refer to it to resolve dependencies on prosodic events that are yet to come. This view was further strengthened by the experiments conducted by [6], which suggest that the look-ahead capacity of speakers might be in fact very long. The precise algorithm for the derivation of prosody from syntax can vary, we show one such algorithm in Fig.1.

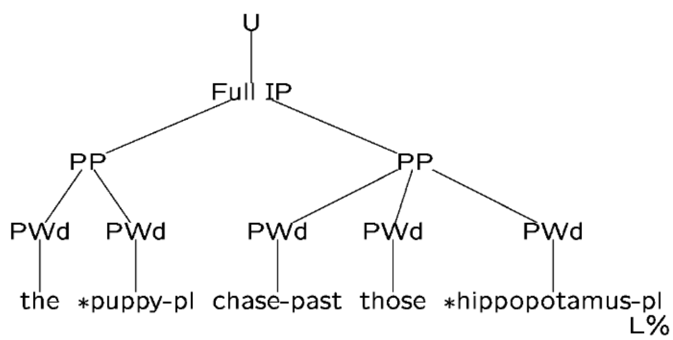

Fig. 2: Prosodic hierarchy tree for the sentence: 'The puppies chased those hypopotamuses.' before restructuring. Adapted from [5]

After the prosodic structure is built, in order to account for other effects, such as eurhythmy, or balanced length of phrases, this default prosodic representation has to be restructured. As an example of restructuring, see Fig.3, which applies a rule that requires phrases to be of roughly the same length.

Interestingly, Jun [4] argued that for some word-form effects, it is not the syntax, but the intonation of the sentence that matters. They demonstrate it by examining two different utterances of the exactly same sentence, 'the black ankle of a cat', in Korean. In one utterance, there is only one intonational phrase, while in the other, there are three. Accordingly, the lenis stop rule is applied differently in both sentences, meaning that, in one utterance, it is pronounced as

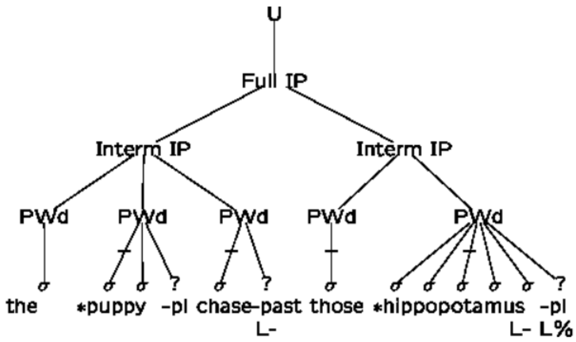

Fig. 3: Prosodic hierarchy tree for the sentence: 'The puppies chased those hypopotamuses.' after restructuring. Note that the word 'chase' was moved to the first phrase. Adapted from [5]

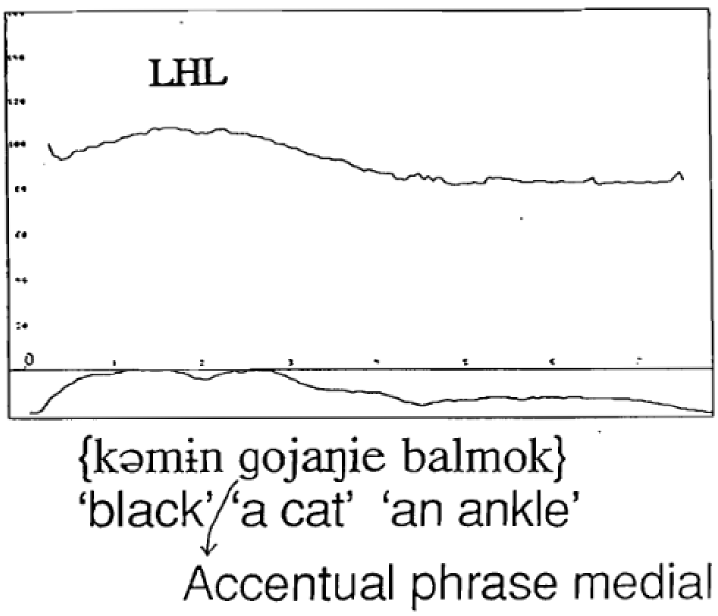

Fig. 4: F0 contour for an utterance of the sentence 'A black cat's ankle.' in Korean. Note that there is only a single accentual phrase, and therefore it is pronounced as 'Gojangie Balmok'. Adapted from [4]

'Gojangie Balmok' while in the other as 'Kojangie Palmok'. Thus, they show that the domain of the lenis stop rule is not the syntax, but intonation, since syntax is the same in both sentences. This intonation-first approach has other benefits, namely that basing word form effects on intonation, instead of prosodical hierarchy derived from syntax, avoids the need for post-hoc restructuring.

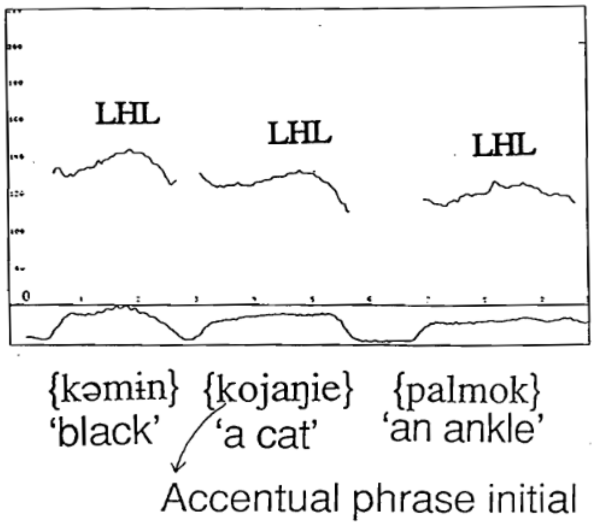

Fig. 5: F0 contour for a different utterance of the same sentence 'A black cat's ankle.' as in Fig. 4. Note that there are three accentual phrases, and therefore it is pronounced as 'Kojangie Palmok', because each both 'Kojangie' and 'Palmok' occur at the beginning of an accentual phrase. Adapted from [4]

3 Prosody in Comprehension

In this section, we will present interesting findings regarding the role prosody plays in comprehension. First, we will present evidence by Beach (1991) [1], who found that speakers can resolve ambiguous sentences by using prosodic cues, even before the whole sentence has been uttered. Then, we will turn to the Implicit Prosody Hypothesis posited by Fodor (2002) [3], which claims that even in silent reading, speakers project prosodic structure onto the read sentence in order to parse it. Fodor provides evidence by showing that prosodic effects can influence how ambiguous sentences are resolved, even in the absence of any overt prosody.

3.1 Complement-verb ambiguities

The beginning parts of utterances can remain ambiguous until a sufficient portion of the utterance is articulated. However, Beach has shown that readers might be able to disambiguate the alternative meanings much sooner. In their experiments, they presented listeners with the beginning of the sentence, asking them to decide which continuation would follow, such as:

(2) Jay believed the gossip ...

(3) Jay believed the gossip about the neighbours right away.

(4) Jay believed the gossip about the neighbours wasn't true.

Listeners were able to decide on the correct relationship even when only hearing the (2) part of the sentence, which has lead Beach to suggest a cue-trading relationship between pitch and duration.

3.2 The Implicit Prosody Hypothesis (IPH)

The Implicit Prosody Hypothesis [2] states that:

> In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction.

The origin of the theory traces back to observing a set of peculiar phenomena, where seemingly unimportant alterations can cause a complete reinterpretation of a sentence. For example, the sentence *I met the* (*recently*) *divorced bishop's daughter.* is parsed differently depending on whether the word 'recently' is present.

(5) I met the divorced bishop's daughter. (I met the daughter of the **divorced bishop**.)

as opposed to

(6) I met the recently divorced bishop's daughter. (I met the bishop's **divorced daughter**.)

A similar effect was noticed in Japanese, with the sentence:

(7) (kyokutanni) shinsetsuna gakusei-no imooto (The (extremely) kind student's daughter)

Interestingly, Fodor [3] notes that a similar effect can be observed even without adding an additional word, but simply by substituting a different noun, as in

(8) I met the niece of the king Spaniards hated. (Spaniards hate the **niece**.)

(9) I met the niece of the king Italians hated. (Italians hated the **king**.)

To explain these effects, Fodor suggests that prosodic effects might be responsible. In examples 8 and 9, Fodor explains that there is a stress clash in version 8, which readers interpret as a structural break, causing them to attach the relative clause to the noun 'niece'. This stress clash is absent in version 9, where the local attachment to 'king' is preferred. This shows that the prosodic stress clash is present, even in silent reading, and can be mistaken for a hint to attach the relative clause non-locally. Therefore, prosody can affect the interpretation of a silently read sentence in the same way as when the sentence is spoken.

As an interesting corollary, it has been found that the same sentence can exhibit different attachment preferences in different languages, such as

(10) Someone shot the servant of the actress who was on the balcony. 60% LC

where readers interpret that 'the actress was on the balcony', as opposed to the Spanish version of the same sentence

(11) Alguien disparó contra la criada de la actriz que estaba en el balcón. 40% LC

where readers tend to interpret that 'the servant was on the balcony'. For a thorough analysis of similar effects in a Slavonic language (Croatian), [8] contains a number of experiments with sentences such as:

(12) Pokusali smo nazvati kcerku politicarke koja/sto voli pjevati u drustvu. (We tried to phone the daughter of the politician who likes to sing when in company)

and similar.

Conclusion

The evidence reviewed in this paper suggests that the role of prosody is not that of an embellishment carrying supplemental information, but rather that it permeates many different aspects of language, and is playing an integral role both in speech planning and comprehension. Even when prosody is absent, evidence points to speakers projecting prosody in order to parse the meaning of a sentence, suggesting that it is deeply intertwined with the underlying cognitive mechanisms. We hope that, in the future, the scientific community will focus more of its attention to these effects, as we believe that many essential truths about language and the human mind can be uncovered by treating prosody as a fundamental, rather than complemental feature.

References

1. Beach, C.M.: The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. Journal of Memory and Language **30**(6), 644–663 (1991). https://doi.org/https://doi.org/10.1016/0749-596X(91)90030-N, https://www.sciencedirect.com/science/article/pii/0749596X9190030N
2. Fodor, J.D.: Prosodic disambiguation in silent reading. In: Proceedings of the 32nd North East Linguistic Society (NELS 32). pp. 113–132. GLSA Publications (2002)
3. Fodor, J.D.: Psycholinguistics cannot escape prosody. In: Proceedings of Speech Prosody 2002 (1st International Conference on Speech Prosody). ISCA (April 2002). https://doi.org/10.21437/SpeechProsody.2002-12
4. Jun, S.A.: The Phonetics and Phonology of Korean Prosody. Phd dissertation, The Ohio State University, Columbus, OH (1993)

5. Keating, P.A., Shattuck-Hufnagel, S.: A prosodic view of word form encoding for speech production. Working Papers in Phonetics 101, UCLA Phonetics Laboratory (2002), https://linguistics.ucla.edu/people/keating/pakssh.pdf
6. Krivokapić, J.: Speech planning and prosodic phrase length. In: Proceedings of Speech Prosody 2010. Chicago, IL, USA (2010). https://doi.org/10.21437/SpeechProsody.2010-264, https://www.isca-speech.org/archive/SpeechProsody_2010/krivokapic10_speechprosody.pdf
7. Levelt, W.J.M., Roelofs, A., Meyer, A.S.: A theory of lexical access in speech production. Behavioral and Brain Sciences **22**(1), 1–38 (1999). https://doi.org/10.1017/S0140525X99001776
8. Lovric, N.: Implicit Prosody in Silent Reading: Relative Clause Attachment in Croatian. Ph.D. thesis, City University of New York (2003), https://academicworks.cuny.edu/gc_etds/3541